我国科学数据管理平台建设成就、缺失、对策及趋势分析*
——基于国内外比较视角

崔旭 赵希梅 王铮 杜丰瑞
西北大学公共管理学院 西安 710127

摘要：[目的/意义]我国科学数据管理平台已经取得了一些成就，但是与国外对比还存在一些差距。拟通过纵向时间轨迹和横向国内外对比对我国科学数据管理平台发展现状进行把脉和定位，同时为国内平台的优化提供国际经验和对策。[方法/过程]采用网络调查手段，从纵向时间轨迹和横向国内外对比两个视角进行研究：通过纵向时间轨迹分析我国科学数据管理平台取得的成就，通过横向对比分析存在的问题。[结果/结论]①成就：相关政策体系不断完善；数据集覆盖的学科领域日渐增多；数据管理平台建设的国际化发展已初见成效。②缺失：资金来源单一；平台服务功能不全面；部分平台缺乏合作建设理念；高校图书馆参与较少等。③对策：建立多元化的资金投入机制；建立数据管理价值链；加强异质机构之间的合作；拓展平台服务方式；高校图书馆应成为数据管理的重要力量。④发展趋势：科学数据管理平台建设将会成为科研信息服务机构的一项重要工作；科学数据管理机构和人员数量会不断扩大；通过建立科学数据管理平台联盟和实力雄厚的科学数据中心来提升科学管理的规模与国际竞争实力。

关键词：科学数据 科学数据管理平台 平台成就 平台对比
分类号：G203
DOI：10.13266/j.issn.0252-3116.2019.09.003

当前，科学研究范式发生了巨大变化，在科学研究活动中，越来越多地利用技术工具对各类研究对象开展数据监测、收集和分析活动，产生大量的科学数据成果（数据集），这些数据成果亟需专业化的管理和利用，于是科学数据平台应运而生。科学数据管理平台专指为科研人员提供科学数据服务的网站、数据库、数据中心等服务媒介。通过这一媒介，研究人员上传或获取研究数据，使得科学活动更加高效和低成本，科研过程更加规范，科研产出成果质量更高。

本文对国内所有数据平台做了调查，并将其分成两大类型：在Re3data[1]上注册的数据平台（见表1）和未在Re3data上注册的数据平台，根据调查发现，我国科学数据平台建设取得了一些成就，但也存在一些问题。因此，本文将从纵向时间轨迹和横向国内外对比两个视角进行分析，纵向分析我国科学数据管理平台取得的成就，横向对比缺失之处，并提出对策建议指出未来发展趋势。希望这项研究能够对国内数据平台优化提供国际经验，推动国内平台的建设与发展。

1 国内外文献评述

通过查阅发现，国外关于数据管理平台方面的研究主要有以下几个方面：①关于平台的调查研究。M. Halbert[2]对爱荷华州立大学、堪萨斯州立大学、俄克拉何马州立大学、内布拉斯加大学林肯分校四所校级图书馆网站所开展的数据管理服务进行了调查。②对已有数据管理平台的介绍。A. L. Vaccarino等[3]介绍了加拿大安大略省大脑研究所的“Brain-CODE”数据管理平台，它可以对不同类型的神经学数据进行收集、存储、整合、共享和分析。T. Nind等[4]介绍了一个用于医学临床数据的开源管理平台，不但满足研究者的数据管理需求，还能支持不同研究小组之间的无缝协作、对数据质量进行评估。③对科学数据管理平台的技术

* 本文系西北大学“本科教学质量提升计划”教学研究与成果培育项目“大数据环境下的图书情报与档案管理本科专业改革与优化路径”（项目编号：JX17006）和“信息资源管理类专业嵌入大数据应用能力与课程链设计研究”（项目编号：JX17045）研究成果之一。
作者简介：崔旭（ORCID：0000-0002-2305-2386）教授，博士生导师，E-mail：cu525.student@sina.com；赵希梅（ORCID：0000-0001-6809-8997），硕士研究生；王铮（ORCID：0000-0001-5727-5935），讲师，博士后；杜丰瑞（ORCID：0000-0001-8606-1821），本科生。
收稿日期：2018-05-17 修回日期：2018-12-15 本文起止页码：21-30 本文责任编辑：杜杏叶

研发。A. D'Anca 等[5]提出了一个可以快速存储和导航海洋学数据的高效、安全和可互操作的数据平台解决方案。L. Persoon 等[6]设计了一个由工作流管理模块、数据管理模块和数据存储管理模块构成的用于管理临床前研究数据的数据管理平台。B. Wang 等[7]提出了一个基于语义网搭建计算化学数据管理平台的方案。

国内主要有两个研究方向:一是科学数据管理平台建设方面的研究;二是科学数据管理平台技术研发和应用的研究。在平台建设方面的研究主要有:①对国内外科学数据管理平台的介绍。如张莎莎[8]、赵卫利等[9]介绍了国外多个科学数据管理平台;陈秀娟等则介绍了化学领域科学数据管理平台[10]。②科学数据管理平台建设标准和规范研究。何毅等对国家人口与健康科学数据共享服务平台中的临床决策支持标准进行了研究[11],顾双双则分析了护理学构建数据平台的元数据标准[12]。③科学数据管理平台绩效评价体系的研究。司莉等构建了平台绩效评估体系[13]。

关于平台技术研发和应用的研究有:①关于平台建设技术和系统软件的研究。张计龙等总结了我国平台软件实践可借鉴的经验[14]。②关于平台运行机制和系统模型的研究。孙仙阁、高芹等论述了平台构建的必要性、原则、要素以及难点,介绍了科学数据管理平台建设的流程、原理、架构和模型[15-16];戴琼洁提出陕西省科学数据管理平台运行机制[17];李花安分析了郑州市科学数据共享的数据库系统模型[18]。③平台系统设计和实现研究。方利等、安基文等、马红旺、程渭介绍了环境科学领域的技术选型和架构[19-22],李花安和付傑等对环境领域科学数据管理平台进行了设计[23-24],俞超等对我国基础研究领域科学数据资源服务平台进行了规划设计[25];朱玲等、罗鹏程等分析了北京大学开放研究数据平台的构建[26-27];王智等[28]、徐淑娟等[29]、张彤等[30]分别对大同地区、煤炭行业、高职院校的科学数据服务平台提出了设计构想;王婷婷提出了军队院校图书馆科学数据安全监管平台的设想[31];吴宁博对大学开放应用平台进行了建构[32];刘润达等对网络科学数据资源的分类导航平台进行了构建[33]。

综上所述,在科学数据管理平台研究方面,国内外侧重技术、应用、以及平台建设存在的问题,尚未从纵向角度对国内科学数据管理平台建设成就进行总结,也没有发展趋势的分析,同时,本文中外对比内容更加全面,全方位分析了我国科学数据管理平台建设现状、取得的成就、存在问题、以及对策建议、未来发展趋势。

2 国内外数据管理平台对比分析

2.1 国内外科学数据管理平台资金来源比较

国内科学数据管理平台的经费几乎都来自国家投入,第一,国家基础科学数据共享服务平台,由科技部专项资金投入。第二,机构独立建设的平台,也是全部由公共财政承担,有的由政府专项资金,有的由"211"建设经费、教育部专项资金、国家自然科学基金、高校自身投入等方式进行建设。与国内相比,国外科学数据管理平台的建设经费来源渠道更为多样,除了来自政府财政的支持外,还有来自专业协会、私营部门、个人捐款等渠道的投入,有的通过会员制形式收取资金,资金来源渠道多。

2.2 国内外科学数据管理平台在 Re3data 上注册数量比较

由于 Re3data 是国际公认的各国共同参与的大规模科学数据库共享平台,可以体现出各国平台建设规模和水平,因此本文以 Re3data 为例,比较各国建设情况。截至 2018 年 7 月,一共有 70 个国家、2 个地区(中国香港和中国台湾)以及国际组织在 Re3data 上注册了科学数据库,注册的全球科学数据库一共有 2 960 个。中国大陆地区(除香港和台湾)在 Re3data 上注册的机构有 38 个,全球排名第 11 位。如图 1 所示:

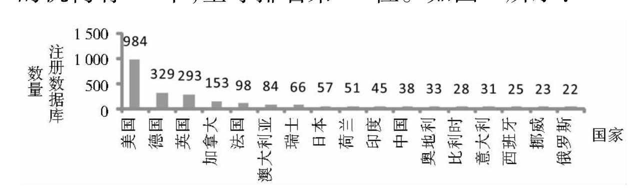


图 1 17 个国家在 Re3data 网站上注册的科学数据库数量统计

笔者根据图 1 的统计结果,将 17 个国家划分为三个梯度。第一梯度(100 个以上数据库):美国、德国、英国、加拿大;第二梯度(50 - 100 个):法国、澳大利亚、瑞士、日本、荷兰;第三梯度(50 个以下):印度、中国、奥地利、比利时、意大利、西班牙、挪威、俄罗斯。笔者重点比较我国与第一梯度国家的差别,美国注册数量是我国的 25 倍左右;德国是我国的 8 倍左右;第一梯度国家注册数量平均值是 439.75 个,是我国的 11 倍左右。可以看出,我国注册数量与第一梯度国家差距悬殊,这说明我国科学数据管理平台的建设在数量上还有很大的发展空间。

崔旭，赵希梅，王铮，等. 我国科学数据管理平台建设成就、缺失、对策及趋势分析——基于国内外比较视角[J]. 图书情报工作,2019,63(9):21－30.

2.3 国内外科学数据管理平台代表性软件开发模式比较

科学数据管理平台系统建设方式有自主研发(含合作研发)、商业软件、开源、委托第三方平台。笔者选取了国内外数据平台影响力较大、软件开发较具典型性的一些平台进行分析和比较(见表1),比如美国自主开发的 Dataverse、Dspace,英国和挪威合作开发的 Nesstar 都是非常有影响力的开源软件。

国内平台软件建设模式大致如下:①自主研发模式。包括:第一,集数据管理和科研团队共享协作为一体的数据管理技术体系,如中国科学院与多家单位合作研发的"数据管理与服务技术体系",包括有科研团队数据管理工具 TeamDR、数据自主管理与发布工具 VisualDB、数据服务注册系统 RSR、可视化服务平台 DVIZ 等 20 余项软件工具[34]。第二,只有数据管理功能的平台,如全国农村固定观察点调查管理和数据分析平台、国家卫计委流动人口数据平台、土地调查成果共享应用服务平台、华大基因等。②采用开源软件开发模式,如北京大学开放数据平台、武汉大学高校科学数据共享平台等。

对于国外数据平台,本文选取了几个典型平台,以自主研发为主,有的是拥有专业的技术研发团队,技术成熟,开发数据管理平台各技术模块,比如哈佛大学—麻省理工学院数据中心开发的 Dataverse 系统;有的是与知名研发机构合作开发,比如美国麻省理工学院图书馆和美国惠普公司实验室合作开发的 DSpace 系统。

表1 国内外科学数据管理平台代表性软件开发模式比照

	国内平台名称	平台软件	国外平台名称	平台软件
在 Re3data 上注册的平台①	国家基础科学数据共享服务平台和各分学科平台② http://www.nsdata.cn/	自主研发	美国密歇根大学校际政治和社会科学研究联盟(ICPSR) https://www.icpsr.umich.edu/icpsrweb/	自主研发
	北京大学开放研究数据平台 http://opendata.pku.edu.cn/	开源 Dataverse	美国哈佛大学—麻省理工学院数据中心(HMDC) https://dataverse.harvard.edu/	自主研发 Dataverse
	华大基因 http://www.genomics.cn/	自主研发	美国国家冰雪数据中心(NSIDC) https://nsidc.org/data	自主研发
未在 Re3data 上注册的平台	清华大学中国经济社会数据研究中心③	由国家统计局自主研发	美国国家生物技术信息中心(NCBI) https://www.ncbi.nlm.nih.gov/	自主研发平台系统及生物数据搜索软件 Entrez
	中山大学学术研究数据库共享计划④	开源 Dspace		
	中国人民大学中国国家调查数据库 http://cnsda.ruc.edu.cn/index.php	自建 SDA	美国地球观测数据网(DataOne) https://www.dataone.org/	自主研发
	华中科技大学中国高校社会科学数据中心 https://cmis.csdc.info/toIndex.action	－	美国康奈尔大学罗普公众舆论研究中心(ROPER) https://ropercenter.cornell.edu/about-the-center/data-curation/	自主研发 RoperExpress
	武汉大学高校科学数据共享平台 http://www.lib.whu.edu.cn/kxsj/aboutus.htm	开源 Dspace	美国芝加哥大学全国民意研究中心的综合社会调查项目(GSS) https://gssdataexplorer.norc.org/	自主研发 Data Explorer
	复旦大学社会科学数据平台 https://dvn.fudan.edu.cn/home/	开源 Dataverse	美国劳工统计局的全国跟踪调查(NLS) https://nsidc.org/data	自主研发
	湖南大学经济数据研究中心 http://edrc.hnu.edu.cn/	－	宾州州立大学图书馆科学数据服务 https://libraries.psu.edu/research/research-data-services	自主研发 Scholar Sphere
	同济大学科研数据管理与服务平台⑤	开源 CHAIR,DSpace	英国数据档案馆(UDA) http://data-archive.ac.uk/	与挪威研究数据中心(NCRD)合作研发 Nesstar
	上海外国语大学数据学术服务平台 https://datam.shisu.edu.cn/home/	开源 Dataverse	澳大利亚国家数据服务网(ANDS) https://www.ands.org.au/working-with-data	自主研发
	全国农村固定观察点调查管理和数据分析系统⑥	自主研发	弗吉尼亚大学图书馆科学数据服务平台 https://data.library.virginia.edu/	与康奈尔大学合作研发 Fedora
	国家卫计委流动人口数据平台 http://www.chinaldrk.org.cn/wjw/#/home	自主研发	美国麻省理工学院图书馆 https://libraries.mit.edu/	与惠普实验室合作研发 Dspace
	土地调查成果共享应用服务平台 http://tddc.mlr.gov.cn/to_Login	自主研发		

注:①Re3data 网址:https://www.re3data.org/browse/by-country/;②国家基础科学数据共享服务平台包括 15 个分学科平台(不包括文献、仪器等保障性平台),这些分平台已在 Re3data 上注册(见表2);③清华大学中国经济社会数据研究中心数据服务功能外网无法访问,没有访问网址;④⑤⑥情况同③

2.4 国内外科学数据管理平台服务功能比较

笔者认为,平台服务功能可分为核心服务功能和其他服务功能,核心服务功能包括数据管理计划、数据创建、数据存储、数据获取、数据分析、数据共享;其他服务功能包括用户指南、用户培训。

笔者调查发现(见图 2),国外数据平台无论是核心服务功能还是其他服务功能都较齐备,在核心服务功能上,都有数据管理计划这一环节,即在网站上设置 Data Management Planning/Data Management Planning Support 链接,告知用户在开展科学研究之前,应当先做好数据管理计划,主要内容包括:数据用户应了解数据创建、存储、获取、分析、共享全过程,在具体实施过程中都应认真完成;知晓涉及科研数据和项目合作者、投资方的相关法律规定和科研道德约束;要将数据管理作为研究中的一个组成要素,不要将数据管理与科研项目目割裂开;审查整个研究中的数据管理,是否有纰漏和错误之处,接受数据管理和共享的培训等。在其他服务功能方面,注重培训方式的多样化,满足用户多方需求。如哈佛大学平台开展了科学数据管理讲座和培训服务,以专题讨论(Workshop)、讲座(Lectures)和在线学习(Online)等方式进行。又比如,弗吉尼亚大学图书馆可以提供一对一的咨询服务。另外,多个平台都将自己的咨询服务人员详细联系方式挂在网上。

国内平台建设方面:①一部分平台在数据管理主要环节都具备,但都缺失数据管理计划这一环节。②用户培训方面,培训方式较为单一,多以讲座报告的形式进行培训,专题讨论、讲座、在线自学三种培训方式兼具的较少,个别平台提供咨询服务。

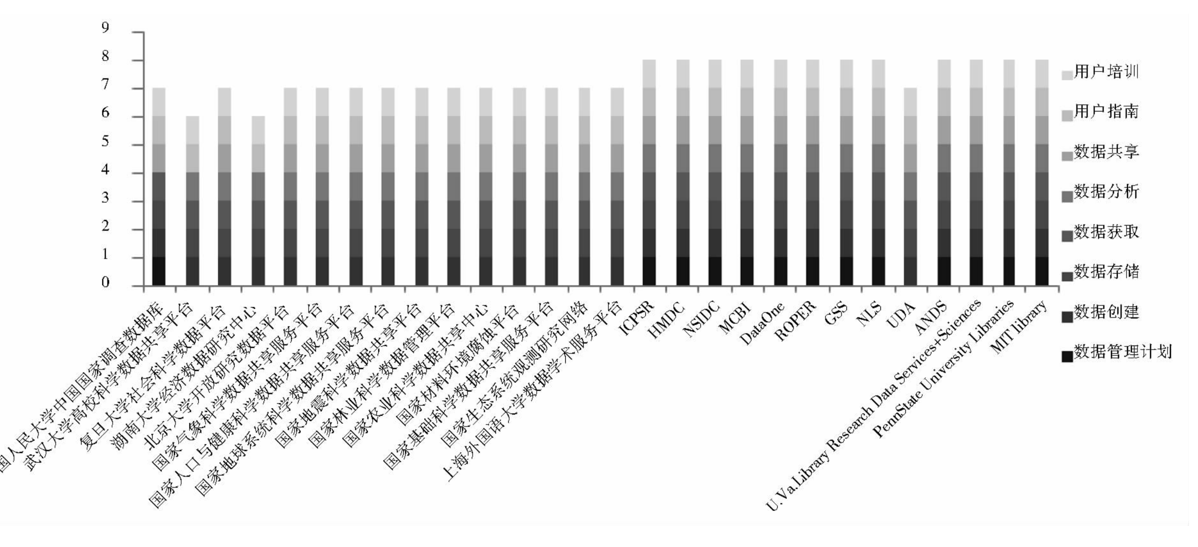


图 2 国内外科学数据管理平台服务功能比较

注:清华大学中国经济社会数据研究中心、中山大学学术研究数据库共享计划、同济大学科研数据管理与服务平台、全国农村固定观察点调查管理和数据分析系统无法打开网站,华大基因不显示提供数据管理服务业务的页面,只显示公司的文字介绍,因此图 2 未做统计。图中平台所有网址链接见表 1

2.5 国内外科学数据管理平台开放共享比较

目前我国科学数据管理平台共享模式有以下几种类型:从共享范围看,有国际范围数据共享、国内范围数据共享、机构范围内数据共享。从共享的组织形式看,有会员制共享模式、注册式共享模式和资格审查制共享模式。①从共享范围看,国际范围数据共享模式。例如在 Re3data 网站上注册的 38 个数据库,都可以为全世界同行公开使用。当然,国内也有未在 Re3data 网站上注册,但与其他国际组织共享数据的平台,比如中国人民大学代表中国加入了国际社会调查合作组织 ISSP(International Social Survey Programme),同时还是 ICPSR(Inter-university Consortium for Political and Social Research)的中国节点。国内范围数据共享是对国内用户开放的一种共享模式,比如农村固定观察点调查系统。机构内共享模式是只对本机构人员提供服务,比如:武汉大学高校科学数据共享平台、湖南大学经济数据研究中心、上海外国语大学数据学术服务平台等。②从共享组织形式看,第一,会员制共享模式,这也是国际上普遍采用的模式。Re3data 网站上注册的国内 38 个数据库所依托的平台管理机构多以会员身份加入 WDC 或者本学科的国际数据共享组织,实现国际数据共享。第二,注册式共享模式。这是国内数据管理平台普遍采取的方式,比如,中国人民大学就通过用户注册实现数据共享。第三,资格审查制共享模式。它

需要数据使用者提交相关申请表，经数据管理部门审核通过后方能获取数据。比如，国家卫计委流动人口数据平台、农村固定观察点调查系统。第四，主动公开共享模式。这是通过互联网主动向公众公开数据。比如土地调查成果共享应用服务平台将脱敏数据主动挂在网站上，供公众使用。

从总体上看，首先，会员制共享是国际上普遍流行的模式，像美国、德国、英国、加拿大这些数据库共享大国，多以会员制身份参与国际数据共享组织活动，并且一些国际数据共享组织都是他们组织发起的。反观国内，虽然有十几家平台参与了国际数据共享合作组织，但是从整体上看，数量还是非常少。第二，各数据共享平台提供的服务功能尚不齐全，有些数据库的服务对象还受到限制。第三，我们尚未建立一个由我国发起的国际数据共享组织，这也是我们未来需要弥补的地方。

3 我国科学数据管理平台建设成就及缺失

3.1 我国科学数据管理平台建设成就

我国科学数据管理平台建设，可以分为几个发展阶段：自发模式阶段(20 世纪 70 年代－2001 年)，以研究所和课题组自主自治为主，但多为独立建设较少共享。数据管理整合阶段(2001 年底 －2012 年)：以 2001 年底科学数据共享工程启动为标志，进入了数据管理整合共享阶段。科学数据国际共享阶段(2013 年－)，以在 Re3data 网站上注册为标志进入了科学数据国际共享阶段，纵观 40 年的发展，在政策法规建设、数据集学科覆盖面、国际化发展方面取得了一定的成就。

(1)数据管理平台建设政策体系在不断完善中。从国家层面到组织层面，我国颁布了相关政策和标准，包括《促进大数据发展行动纲要》《科学数据管理办法》《国家科技资源共享服务平台管理办法》《关于逐步开发应用微观调查数据的试行办法》。从组织层面看，从 2004 年起，先后在基础科学、农业、林业、海洋、气象、地震、地球系统科学、人口与健康数据平台建立了一系列准则和标准，比如农业科学数据共享平台有 38 项技术规范和管理制度，国家材料科学数据共享平台有 23 项。这些为数据加工整理、数据汇交和共享都提供了保障，使得数据管理平台运行更加规范化和科学化。

(2)数据管理平台覆盖学科领域日渐增多。最先开展数据管理与共享工作的是科技部建立的基础科学数据共享服务平台，截止 2017 年底，已聚集参加建设单位 43 家，整合物理、化学、天文、空间与生物领域的 254 个专业数据库，开放共享数据达 431TB，累计访问量 1 712.41 万人次，页面访问量累计 1 亿次，数据下载总量 243.94TB[35]。2009 年开始，我国高校图书馆建立数据管理平台，侧重于社会科学数据的收集。2016 年，政府微观数据开放也提到了议事日程，学科涉及范围越来越广。

(3)数据管理平台建设的国际化发展已初见成效。第一，我国在生物学、海洋、气象、地震、地质、地球物理、空间、天文、冰川冻土、可再生资源与环境学科等领域加入了世界数据中心(World Data Center, WDC)或者所属领域的国际数据合作组织，紧跟国际科学数据管理发展步伐。同时，我国是 WDC 五个地区中心中的一个中心，实现了与世界同行的数据共享。第二，数据源贡献单位往往是多家，不但有国内机构，还有国际研究机构。比如棉花功能基因组学数据库虽然在 Re3data 网站上注册的是一个单位，但数据源实际贡献单位却包括国内、国外的研究组织和数据中心。

3.2 我国科学数据管理平台建设中的缺失之处

虽然我国数据管理平台建设取得了一些成就，但从全国各类型平台来看，还存在着一些问题。

3.2.1 资金来源单一，造成数据管理平台不能均衡发展

科学数据管理平台建设和运行需要持续大量的资金支撑，如果投入不足会出现数据更新困难，平台系统难以维护，服务不佳等问题。通过国内外对比发现，国内科学数据管理平台资金来源单一，几乎全部是各部委专项资金投入，比如国家科技基础条件平台是科技部专项资金投入，有些数据平台是教育部专项投入，比如华中科技大学中国高校社会科学数据中心，有的是国家自然科学基金委，比如北京大学等，有的是“211”建设经费，比如武汉大学，捐赠资金几乎没有，只有清华大学中国经济社会数据研究中心 2018 年获得中华思源工程扶贫基金会闽善公益基金 2 000 万捐款。资金投入渠道的单一化，加重了政府财政负担，政府资金仅能投入一些大型数据平台，投入少的平台其后期维护和建设会受到影响，使得国内数据管理平台发展愈加不均衡。

3.2.2 软件开发缺少开源理念，提高了我国整体平台系统开发成本

我国高校建立的平台都是在开源软件基础上的二次开发，但在汉化过程中，尚存在一些问题，比如，前期使用的某国外开源软件没有数字持久标识符标记功能，还需要研究团队使用别的开源软件嵌

表 2 Re3data 网站上注册的国内科学数据库统计

序号	数据库/平台名称	学科领域	注册建设单位
1	全基因组数据库	遗传学、生物学、医学	华大基因
2	组学原始数据归档库	生物学、遗传学、解剖学、动物学、基础生物学、医学	中国科学院北京基因组研究所
3	MiCroKitS	生命科学、生物学、遗传学、基础生物学和医学研究、植物学、动物学	中国科学技术大学生命科学学院、中国科学技术大学合肥微尺度物质科学国家实验室、The Cuckoo Work-group
4	空间科学虚拟观测站	物理、天体物理学和天文学	中科院中国空间科学数据中心
5	基因组尺度代谢网络模型数据库	生物化学、病毒学、免疫学	天津大学
6	哺乳动物动态转录本数据库	人、猪、大鼠和小鼠转录组数据	中国科学院北京基因组研究所、国家自然科学基金委
7	国际大洋发现计划	地球科学(包括地理)、海洋学、地质学和古生物学、大气科学	中华人民共和国科学技术部
8	数据堂	人文社会科学、生命科学、自然科学	数据堂科技有限公司
9	国家地球系统科学数据共享服务平台	地球科学(包括地理)、地质学和古生物学	中国科学院地理科学与资源研究所
10	北京大学开放研究数据平台	人文社会科学、工程科学、计算机科学、电气与系统工程	北京大学图书馆、北京大学
11	国际小鼠表型分析联盟	动物遗传学	南京大学－南京生物医药研究院
12	家蚕基因组数据库	家蚕基因组数据	西南大学家蚕基因组生物学国家重点实验室
13	家蚕病原数据库	家蚕病原数据	西南大学家蚕基因组生物学国家重点实验室
14	犬类 SNP 数据库	犬类基因组数据	中国科学院北京基因组研究所、中国科学院昆明动物研究所
20	非编码 RNA 组科学数据库	一般遗传学、生物化学、基础生物学和医学研究、生命科学	中国科学院计算技术研究所生物信息学研究组、中国科学院生物物理研究所生物信息学实验室
21	蛋白质赖氨酸修饰数据库	生物学、遗传学、基础生物学和医学研究、植物学、动物学	华中科技大学生命科学与技术学院生物医学工程系、中国科学技术大学生命科学学院、中国科学技术大学合肥微尺度物质科学国家实验室、The Cuckoo Workgroup
22	中国天文数据中心	物理、天体物理学和天文学	中国科学院国家天文台、中国虚拟天文台、科技部、国家自然科学基金
23	动物数量性状基因座数据库	动物学、动物遗传学，细胞和发育生物学、生物学	华中农业大学、美国国家动物基因组研究计划、生物信息学协调计划、爱荷华州立大学、美国农业部国家粮食和农业研究所
24	全球微生物目录	微生物学、病毒学和免疫学、医学、生物学	中国科学院微生物研究所网络信息中心
25	中国南北极数据中心	大气科学与海洋学、地球科学(包括地理)、地球物理学和大地测量学	国家海洋局、中国极地研究所、科技部、国家科技基础条件平台中心、中国科学院地理科学与资源研究所
26	高能物理文献数据库	物理、粒子、核和场	中国科学院高能物理研究所
27	兰州寒区旱区科学数据中心	地球科学(包括地理)	中国科学院寒区旱区环境与工程研究所
28	中国林业科学数据中心	植物科学、生物学	中国林业科学研究院
29	可再生资源与环境世界数据中心	工程科学、物理、地球科学(包括地理)、地球物理学、大地测量学、热能技术、热机、流体力学、天体物理学和天文学	中国科学院地理科学与资源研究所
30	国家气象科学数据共享服务平台	大地测量学、地球科学(包括地理学)、大气科学和海洋学	中国气象局、国家气象信息中心、国家科技基础条件平台中心
31	棉花功能基因组学数据库	植物科学、生物学	中国农业科学院生物技术研究所
32	世界数据中心－海洋学、天津	大气科学与海洋学、地球科学(包括地理)、水研究、动物学、生物学	国家海洋信息中心、国家海洋局
33	GigaDB	基础生物学、遗传学、植物科学、动物学、微生物学、病毒学、免疫学、医学	华大基因、GigaScience 杂志社、国家基因库大鹏总部

（续表2）

序号	数据库/平台名称	学科领域	注册建设单位	序号	数据库/平台名称	学科领域	注册建设单位
15	全球变化研究数据发布和存储库	大气科学、海洋学、地球科学(包括地理)，大地测量学、遥感学	中国地理学会、中国科学院地理科学与资源研究所	34	世界鱼类数据库	动物学、动物生态学、动物遗传学、生物学	中国水产科学研究院
16	必须基因数据库	生物学和医学研究、遗传学、病毒学和免疫学、生命科学	天津大学生物信息中心	35	国家数据库	统计数据	国家统计局
17	中国作物种质资源信息系统	植物学、生物学、生命科学	中国农业科学院作物科学研究所	36	世界微生物数据中心	微生物学、病毒学、免疫学、医学、生物学	中国科学院微生物研究所网络信息中心
18	国家地震科学数据共享中心	大气科学与海洋学、地球科学(包括地理)、地球物理学和大地测量学	中国和国际组织共建	37	世界数据中心中国地球物理数据中心	地球物理学和大地测量学、大气科学与海洋学、水研究、地质学和古生物学、地球科学(包括地理)	中国科学院地质与地球物理研究所
19	植物转录因子数据库	生物信息学与理论研究、生物学、植物遗传学、基础生物学和医学研究	北京大学生物信息中心、国家863项目、973项目、国家自然科学基金委、教育部	38	原子分子数据库	光学、量子光学和原子、分子和等离子体物理学、粒子、核和场、物理	北京应用物理与计算数学研究所、国际科学技术数据委员会(CODATA)中国全国委员会

注：表中38个数据库网址链接：https://www.re3data.org/search?query=&countries[]=CHN

入该功能；国内各汉化研究团队之间缺乏交流，使得汉化过程出现的问题不能及时解决。第二，国家科学数据共享工程、政府部门、商业机构建设的平台以自主研发居多，他们有技术研发团队，比如中科院计算机网络信息中心、华大基因、国家各部委信息中心等。但是，我国即使是自主开发，也只是仅供本系统内使用，没有开发出像Dataverse系统、DSpace系统这样具有世界影响力的开源软件，缺乏开源思想和共享理念，由于国内开发的系统不是开源系统，成效只能在本系统内显现，不能为国内其他机构所使用，提高了我国整体平台系统开发成本。

3.2.3　部分平台存在单打独斗现象，缺乏合作建设理念　国外科学数据管理平台合作建设情况较多，比如校际政治和社会科学研究联盟(ICPSR)，它由12所高校、1家政府机构(联邦预算管理局)组成理事会，会员单位有776家，联合多个机构进行合作建设，形成了非常高的世界影响力。国内虽然以中科院为建设主体的数据平台合作单位多，数据资源较丰富，但未进入国家科技基础条件平台的高校，则合作建设较少，单打独斗、不成规模，整体来看，缺乏不同类型机构之间的合作交流，比如高校和研究所之间，研究所与企业之间，从国家层面上来讲，缺乏合作必然会影响平台数据规模以及数据共享，不利于国家整体科学数据资源的利用与开发，影响国家整体科技竞争力的提升。

3.2.4　平台服务功能不全面，会影响数据使用的用户黏度　目前，一些科学数据平台提供的服务功能不全面，数据管理计划缺失，用户培训形式单一；一些平台数据下载有较多的限制，服务对象仅限于本单位人员；一些平台名称上冠以数据管理平台，但实际并不提供数据集服务，仅提供文献收藏和服务功能；一些平台数据资源规模小、数据集不连续，影响数据分析质量。服务功能的不完备，直接影响了用户的使用体验，会影响用户使用黏度，数据平台对支撑用户科研的作用弱化。

3.2.5　高校图书馆科学数据管理平台数量少，尚未形成规模　通过调查笔者发现，欧美国家高校图书馆在建设科学数据管理平台有两个特点：一是积极性高，参与的高校馆数量多；二是与其他部门合作建设广泛，高校馆科学数据管理平台已形成规模。相比较而言，我国高校图书馆科学数据管理平台建设数量少，从参与到中国高校图书馆研究数据管理推进工作组的高校馆数量就可以看出来，不到10家，说明我国高校图书馆没有成为科学数据管理平台的生力军。

4　我国科学数据管理平台建设对策

4.1　拓展渠道来源，建立多元化的资金投入机制

科学数据管理平台的运转需要人才资源、技术资源、内容资源的支撑，但随着数据服务规模的扩大，平台运转所需资金量也会越来越大，仅依赖国家财政的

支持,这是不够的,需要构建多元化的资金来源渠道,为平台建设提供更多支持。因此,应当借鉴国外经验,比如哈佛 - 麻省理工社会科学数据中心(HarvardMIT Data Center,HMDC),1999 年,HMDC 获得美国国家科学基金会和其他 5 家资助机构数百万美元的赠款,此后,HMDC 从美国国会图书馆等处获得额外的补助和资金支持[36],既鼓励地方政府、高等院校、科学基金对平台建设的支持,还应制定优惠政策积极吸引社会组织、企业和个人对平台的资助,建立多元化的资金投入机制。另外,也可以通过会员制形式,获取资金支持。

4.2 引导国内 IT 企业和非营利组织开发数据管理软件,降低国家整体开发成本

目前,从国外发展现状看,数据服务已经是一项重要服务项目,可以预见,未来国内会有越来越多的科研院所开展数据管理服务,但并不是每个机构都有技术实力自主开发软件或者汉化国外软件,因此,为了节省成本,减少重复投入和重复建设,国家应制定相关政策,鼓励国内 IT 企业开发数据管理软件,同时,也鼓励非营利组织开发开源软件,为国内各科研院所使用,降低国家整体开发成本,推动数据服务的普及。

4.3 加强异质机构之间的合作,创新数据共建共享模式

借鉴美国地球观测数据网(DataONE)共建模式,发展多个协调机构和成员机构,与图书馆、数据库商、赞助商等建立合作关系,我国科学数据管理平台应加强异质机构之间的合作,科研机构、高校、企业构成科学数据管理协作链条,分工协作、发挥优势、形成体系,使科研人员不会因为所在单位的数据资源局限而影响研究质量,借助数据共建模式多渠道获取数据资源。共建数据平台共享科学数据是未来发展的主要趋势,只有平台的数据集形成规模才能与世界各国的数据平台交流共享数据资源。

4.4 完善平台服务功能,建立为科研服务的数据管理价值链

我国科学数据管理平台应学习和借鉴国外科学数据管理平台服务功能全的经验,完善我国数据管理平台服务功能,在平台上增设数据管理计划功能,引导用户在科研活动中形成良好的数据规划行为;扩充咨询服务人员队伍数量,增加培训方式和手段,比如一对一咨询、专题讨论、讲座、在线自学等,服务功能全面,才能提高数据用户的体验感,增强数据使用的用户黏度,从而形成一个良性的数据管理循环链条,真正建立起为科研服务的数据管理价值链。

4.5 高校图书馆要寻求与其它部门的合作,成为科学数据管理平台建设主力

高校图书馆在技术上要积极与 IT 企业合作,在软件平台开发上保证完备的技术支持,建立的平台服务功能应该齐全,服务途径应该多样,服务方式应该便捷;在数据集建设方面,要积极与校内院系以及校外研究所、学术团体的合作,联合进行数据管理建设。与科研机构的紧密合作是高校图书馆建设数据管理平台、开展数据服务的最佳路径。

5 我国科学数据管理平台发展趋势分析

5.1 趋势一:数据管理平台建设将会成为科研信息服务机构的一项重要工作

未来科学研究工作都越来越依赖数据资源的支撑,集中统一的数据管理可以降低数据整理成本,防止数据的丢失和损坏,避免造假数据的出现,推进科学研究成果整体质量的提高,所以说,国家及研究机构都会越来越重视数据管理工作,它将成为未来高校图书馆、科研图书馆、科研机构、信息服务机构的一项重要工作内容。

5.2 趋势二:科学数据管理机构和人员数量会不断扩大

未来会有越来越多的科学信息服务机构建立数据管理平台,需要更多的专门从事科学数据管理的专业人员,目前,这一领域的人才培养还是非常薄弱,我国只有个别高校设立了与数据管理相关的专业,未来还需要大批这样的专业人才,因此高校相关专业的人才培养将成为未来发展的一个趋势。

5.3 趋势三:通过建立科学数据管理平台联盟来提升科学管理的规模与竞争实力

从国外成功经验可以看出,科学数据管理平台只有形成规模才能促进科学数据的共享与科研活动的开展,因此,未来应当是建立科学数据管理平台建设联盟,不但是高校图书馆之间,还应包括高校图书馆、科研院所图书馆、企业数据管理中心之间,这样才能从根本上解决国内数据集规模和数据集质量问题,为国家的科技创新提供有力支撑。

5.4 趋势四:通过建立实力雄厚的科学数据中心提升国际影响力

我国要向科技强国发展,前提之一是要发展成为数据强国,因此建设数据集质量高、规模大的科学数据中心,是未来的发展方向,目前,我国有个别数据中心是国际数据中心的主持单位,但是这样的主持单位还

是凤毛菱角，因此要通过建立实力雄厚的科学数据中心，让更多的学科数据管理平台成为国际数据中心的主持单位，提高国际影响力，推动向数据强国的目标发展。

参考文献：

[1] re3data. org[EB/OL]. [2018－06－04]. https://www. re3data. org/browse/by-country/.

[2] HALBERT M. Prospects for research data management. Research data management: principles, practices, and prospects[R/OL]. [2018－03－12]. http://libres. uncg. edu/ir/uncg/f/M_Halbert_Prospects_2013. pdf.

[3] VACCARINO A L, DHARSEE M, STROTHER S, et al. Brain-CODE: a secure neuroinformatics platform for management, federation, sharing and analysis of multi-dimensional neuroscience data [J/OL]. Frontiers in neuroinformatics[2018－06－10]. https://doi. org/10. 3389/fninf. 2018. 00028.

[4] NIND T, GALLOWAY J, MCALLISTER G, et al. The research data management platform (RDMP): a novel, process driven, open-source tool for the management of longitudinal cohorts of clinical data[J]. GigaScience, 2018,7(7):1－12.

[5] D'ANCA A, CONTE L, NASSISI P, et al. A multi-service data management platform for scientific oceanographic products[J]. Natural hazards and earth system sciences,2017(17):171－184.

[6] PERSOON L, HOOF S V, KRUIJSSEN F V D, et al. A novel data management platform to improve image-guided precision preclinical biological research[C/OL]. Paper for special issue for 4th Conference on Image guided precision radiotherapy[2018－10－30]. https://www. birpublications. org/doi/10. 1259/bjr. 20180455.

[7] WANG B, DOBOSH P A, CHALK S, et al. Computational chemistry data management platform based on the semantic web[J]. The journal physical chemistry,2017,121(1):298－307.

[8] 张莎莎，黄国彬，耿骞. 基于 re3data 的英国科学数据发布平台研究[J]. 数字图书馆论坛,2017(6):16－24.

[9] 赵卫利，陈晓毅，靳红. 科学数据共享平台，支撑优势产业发展研究[J]. 科技与经济,2008(2):53－55.

[10] 陈秀娟，吴鸣，胡卉. 嵌入科研工作流的图书馆数据管理服务——以化学学科为例[J]. 图书馆论坛,2016(3):49－55,102.

[11] 何毅，王曙光，刘文浩. Infobutton 在国家人口与健康科学数据共享平台的应用研究[J]. 中国数字医学,2016(1):80－83.

[12] 顾双双. 护理学科学数据服务平台的构建[J]. 医学信息学杂志,2014(7):33－36,49.

[13] 司莉，李月婷，邢文明，等. 我国科学数据共享平台绩效评估实证研究[J]. 图书馆理论与实践,2014(9):30－35.

[14] 张计龙，殷沈琴，张用，等. 社会科学数据的共享与服务——以复旦大学社会科学数据共享平台为例[J]. 大学图书馆学报,2015(1):74－79.

[15] 孙仙阁. 云环境下高校图书馆科学数据集成与共享服务平台构建研究[J]. 图书馆学刊,2016(5):133－136.

[16] 高芹，钟晓莉. 高校图书馆科学数据监护平台构建研究[J]. 图书馆学刊,2015(8):113－116.

[17] 戴琼洁. 陕西省科学数据共享平台运行机制研究[D]. 西安：西安电子科技大学,2011.

[18] 李花安，常朝稳. 科学数据共享平台 Web 服务中间层的设计与实现[J]. 微计算机信息,2006(12):247－249.

[19] 方利，王文杰，高振记，等. 基于 SOA 的环境科学数据共享平台设计与实践[J]. 环境工程技术学报,2014(4):333－340.

[20] 安基文，庄大方，袁文. 面向地学计算的资源环境科学数据共享平台的设计[J]. 地球信息科学,2007(3):34－39.

[21] 马红旺. 基于 Geoportal 的环境科学数据共享平台研究与实现[D]. 湘潭：湖南科技大学,2012.

[22] 程渭. 空间环境科学数据共享平台研究与实现[D]. 武汉：中国地质大学,2010.

[23] 李花安. 基于全局数据库的科学数据共享平台的研究与实现[D]. 郑州：解放军信息工程大学,2006.

[24] 付傑. 基于网络地理信息系统(WebGIS)的东海科学数据节点平台的设计与实现[J]. 科技情报开发与经济,2008(14):148－150.

[25] 俞超，蔡维华，洪镇洲. 基于 GIS 的地震科学数据共享平台的设计与实现[J]. 计算机应用与软件,2011(7):86－88,158.

[26] 朱玲，聂华，崔海媛，等. 北京大学开放研究数据平台建设：探索与实践[J]. 图书情报工作,2016,60(4):44－51.

[27] 罗鹏程，朱玲，崔海媛，等. 基于 Dataverse 的北京大学开放研究数据平台建设[J]. 图书情报工作,2016,60(3):52－58.

[28] 王智，王莉. 大同科学数据共享平台设计与实现[J]. 软件导刊,2015,14(12):81－83.

[29] 徐淑娟，江艳霞. 煤炭高校图书馆联盟构建行业科学数据共享平台研究[J]. 情报探索,2016(9):67－70,75.

[30] 张彤，李明佳. 高职院校科研数据管理平台设计研究——以辽宁金融职业学院为例[J]. 辽宁高职学报,2016(9):92－94.

[31] 王婷婷. 基于可信云计算的军队院校图书馆科学数据安全监管平台构建研究[J]. 图书馆学刊,2016(12):101－104.

[32] 吴宁博. 大学开放科研数据范畴与应用平台建构研究[D]. 贵州：贵州财经大学,2016.

[33] 刘润达，彭洁，涂勇. 一种多维关键词与分类关联的科学数据资源分类导航平台构建方案[J]. 现代图书情报技术,2010(9):74－78.

[34] 黎建辉，沈志宏，孟小峰. 科学大数据管理：概念、技术与系统[J]. 计算机研究与发展,2017(2):235－247.

[35] 国家基础科学数据共享服务平台[EB/OL]. [2018－06－05]. http://nsdata. cn/collection/about.

[36] 张计龙，朱勤，殷沈琴. 美国社会科学数据的共享与服务[J]. 大学图书馆学报,2013(5):13－17.

[37] 崔宇红. 机构研究数据管理实践探析：模型、核心服务和优先战略[J]. 情报理论与实践,2017(8):19－22,29.

[38] CHOURASIA A, WONG M, MISHIN D, et al. SeedMe: a scientific data sharing and collaboration platform[J/OL]. [2017－11－

14]. https://www.seedme.org/sites/seedme.org/files/publications/paper-xsede2016-preprint.pdf.

[39] KRAFT A, RAZUM M, POTTHOFF J, et al. The radarproject: a service for research data archival and publication[EB/OL]. [2017-12-16]. https://www.mdpi.com/2220-9964/5/3/28/htm.

[40] HERRICK R, HORTON W, OLSEN T, et al. XNAT central: open sourcing imaging research data[J]. NeuroImage, 2016(1): 1093-1096.

[41] HERZINGER S, GU W, SATAGOPAM V, et al. SmartR: an open-source platform for interactive visual analytics for translational research data[J/OL]. [2017-12-21]. https://www.ncbi.nlm.nih.gov/pubmed/28334291. Doi: 10.1093/bioinformatics/btx137.

[42] HUANG J, ZHANG X C, EISENHAUER G, et al. Scibox: online sharing of scientific data via the cloud[C/OL]. IEEE 28th International Parallel and Distributed Processing Symposium. Phoenix: AZ, 2014: 145-154[2018-03-04]. https://www.cc.gatech.edu/~jhuang95/papers/scibox-ipdps14-slides.pdf.

作者贡献说明:

崔旭:设计研究框架,指导论文写作,撰写论文部分内容,审定修改;

赵希梅:进行数据调查,撰写论文部分内容;

王铮:进行数据调查,审定修改;

杜丰瑞:数据汇总,绘制图表。

Achievements, Defects, Countermeasures and Trends of Chinese Scientific Data Management Platform Construction-based on Domestic and Foreign Comparative Perspectives

Cui Xu Zhao Ximei Wang Zheng Du Fengrui
School of Public Management, Northwest University, Xi'an 710127

Abstract: [**Purpose/significance**] Chinese scientific data management platform has made some achievements, but there are still some gaps in horizontal comparison with foreign countries. Therefore, this paper aims to analyze the current development status through horizontal and vertical comparison, and provide international experience and countermeasures for the optimization of domestic platforms. [**Method/process**] This paper used the network survey method to study from the perspectives of longitudinal time trajectory and horizontal and domestic: Using network survey means, research from two perspectives: vertical time trajectory and horizontal domestic and international comparison: analyzing the achievements of China's scientific data management platform through longitudinal time trajectory, and analyzing the existing problems through horizontal comparison. [**Result/conclusion**] Achievements: The relevant policy system is continuously improved; the subject areas covered by data sets are increasing; the international development of data management platform construction has achieved initial results. Lack: The source of funds is single; the platform service function is not comprehensive; the cooperation construction is lacking; the participation of university library is less. Countermeasures: Diversified capital investment mechanism; establish a data management value chain for scientific research services; strengthen cooperation between heterogeneous institutions; expand platform service methods; university libraries should become an important force in data management. Development trend: The construction of scientific data management platform will be an important content of scientific research information service institutions; the number of scientific data management institutions and personnel will continue to expand; through the establishment of a scientific data management platform alliance and powerful scientific data center to improve the scale of scientific management and international competitiveness.

Keywords: scientific data scientific data management platform platform achievements platform comparison